PATENT APPLICATION BASED ON:

Docket No: 83,706

Inventors: Nathan D. Cahill
Lawrence A. Ray

Attorney: Thomas H. Close

METHOD AND SYSTEM FOR COMPOSITING IMAGES TO PRODUCE A CROPPED IMAGE

Commissioner for Patents
Attn: Box Patent Application
Washington, DC 20231

Express Mail Label No: EL 486 846 887US
Date: December 19, 2001

METHOD AND SYSTEM FOR COMPOSITING IMAGES TO PRODUCE A CROPPED IMAGE

FIELD OF THE INVENTION

The invention relates generally to the field of digital image processing, and in particular to a technique for compositing multiple images into a large field of view image, said image being cropped to a selected aspect ratio.

BACKGROUND OF THE INVENTION

Conventional systems for generating images comprising a large field of view of a scene from a plurality of images generally have two steps: (1) an image capture step, where the plurality of images of a scene are captured with overlapping pixel regions; and (2) an image combining step, where the captured images are digitally processed and blended to form a composite digital image.

In some of these systems, images are captured about a common rear nodal point. For example, in USSN 09/224,547, filed December 31, 1998 by May et. al., overlapping images are captured by a digital camera that rotates on a tripod, thus ensuring that each image is captured with the same rear nodal point lying on the axis of rotation of the tripod.

In other systems, the capture constraint is weakened so that the images can be captured from substantially similar viewpoints. One example of a weakly-constrained system is the image mosaic construction system described in US Patent 6,097,854 by Szeliski et al., issued August 1, 2000; also described in Shum et al., "Systems and Experiment Paper: Construction of Panoramic Image Mosaics with Global and Local Alignment," IJCV 36(2), pp. 101-130, 2000. Another example is the "stitch assist" mode in the Canon PowerShot series of digital cameras (see http://www.powershot.com/powershot2/a20_a10/press.html;; US Patent 6,243,103 issued June 5, 2001 to Takiguchi et al.; and US Patent 5,138,460 issued August 11, 1992 to Egawa.

In some systems, the capture constraint is removed altogether, and the images are captured at a variety of different locations. For example, the view morphing technique described in Seitz and Dyer, "View Morphing," SIGGRAPH

'96, in *Computer Graphics*, pp. 21-30, 1996, is capable of generating a composite image from two images of an object captured from different locations.

The digital processing required in the image combining step depends on the camera locations of the captured images. When the rear nodal point is exactly the same, the image combining step comprises three stages: (1) a warping stage, where the images are geometrically warped onto a cylinder, sphere, or any geometric surface suitable for viewing; (2) an image alignment stage, where the warped images are aligned by a process such as phase correlation (Kuglin, et al., "The Phase Correlation Image Alignment Method," *Proc. 1975 International Conference on Cybernetics and Society*, 1975, pp. 163-165), or cross correlation (textbook: Gonzalez, et al., *Digital Image Processing*, Addison-Wesley, 1992); and (3) a blending stage, where the aligned warped images are blended together to form the composite image. The blending stage can use a simple feathering technique that uses a weighted average of the images in the overlap regions, and it can utilize a linear exposure transform (as described in USSN: _______, filed November 5, 2001 by Cahill et al., our docket no. 83516/THC) to align the exposure values of overlapping images. In addition, a radial exposure transform (as described in USSN ______, filed December 17, 201 by Cahill et al., our docket 83512/THC) can be used in the blending stage to compensate for light falloff.

In weakly-constrained systems, the image combining step generally comprises two stages: (1) an image alignment stage, where the images are locally and/or globally aligned according to some model (such as a translational, rotational, affine, or projective model); and (2) a blending stage, where the aligned images are blended together to form a texture map or composite image. The blending stage typically incorporates a de-ghosting technique that locally warps images to minimize "ghost" images, or areas in the overlapping regions where objects are slightly misaligned due to motion parallax. The local warping used by the de-ghosting technique can also be incorporated in the model of the image alignment stage. For an example of image combining with such a system, see the aforementioned Shum and Szeliski references.

In systems where the capture constraint is removed altogether, the image combining step first requires that the epipolar geometry of the captured images be estimated (for a description of estimating epipolar geometry, see Zhang, et al., "A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry," INRIA Report No. 2273, May 1994, pp. 1-38). Once the epipolar geometry has been estimated, the images are projected to simulate capture onto parallel image planes. The projected images are then morphed by a standard image morphing procedure (see Beier et al., "Feature-Based Image Metamorphosis," SIGGRAPH '92, *Computer Graphics*, Vol. 26, No. 2, July 1992, pp. 35-42), and the morphed image is reprojected to a chosen view point to form the composite image. An example of such a system is described in the aforementioned Seitz and Dyer reference.

In all of the prior art methods and systems for generating large field of view images, the composite image is provided as output. In some instances, however, it might be necessary to provide a composite image that has been cropped and/or zoomed to a selected aspect ratio and size. For example, consider a digital photofinishing system that prints hardcopies of images that have been digitized from film after being captured by an Advanced Photo System (APS) camera. APS cameras provide the photographer the choice of receiving prints in three different formats: HDTV (H), Classic (C), or Panoramic (P). The Classic format corresponds to a 3:2 aspect ratio, the HDTV format to a 16:9 aspect ratio, and the Panoramic format to a 3:1 aspect ratio. If the photographer captures a sequence of images with an APS camera and uses one of the known techniques to generate a composite image, the composite image will likely not have an aspect ratio corresponding to the H, C, or P formats. Since one of these three formats would be required in the digital photofinishing system, the photographer must manually intervene and crop the composite image to the appropriate aspect ratio for printing.

There is a need therefore for an improved method that will combine images into a composite image; the method being capable of automatically cropping the composite image to a desired aspect ratio.

SUMMARY OF THE INVENTION

The need is met according to the present invention by providing a method for producing a cropped digital image that includes the steps of: providing a plurality of partially overlapping source digital images; providing a cropping aspect ratio L:H, the cropping aspect ratio being the ratio of the length to the height of the cropped digital image; providing a cropping criterion, the cropping criterion being a criterion for the size and location of the cropped digital image; combining the source digital images to form a composite digital image; selecting the cropping region of the composite digital image according to the cropping criterion, said cropping region being a rectangular region having aspect ratio L:H, and having size and location determined by the cropping criterion; and, cropping the composite digital image to the cropping region to form a cropped digital image.

ADVANTAGES

The present invention has the advantage of automatically producing a cropped digital image in a system for compositing a plurality of source digital images. This eliminates the need for the user to crop and/or resize the composite digital image.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram illustrating a digital image processing system suitable for practicing the present invention;

Fig. 2 illustrates in block diagram form, the method of forming a cropped digital image from at least two source digital images;

Fig. 3 illustrates the preferred cropping criterion;

Fig. 4 illustrates an alternative cropping criterion;

Fig. 5 illustrates a further alternative cropping criterion;

Fig. 6 illustrates in block diagram form, an embodiment of the step of selecting a cropping region according to the preferred cropping criterion;

Fig. 7 illustrates in block diagram form, a further embodiment of the step of providing source digital images;

Figs. 8A and 8B illustrate in block diagram form, further embodiments of the step of providing source digital images;

Fig. 9 illustrates in block diagram form, a further embodiment of the step of combining source digital images;

Fig. 10 is a diagram useful in describing the step of combining the adjusted source digital images;

Fig. 11 illustrates in block diagram form, a further embodiment of the step of blending warped digital images;

Figs. 12A and 12B are diagrams useful in describing the aspect ratio of an image; and

Figs. 13A and 13B illustrate a source digital image file containing image data and meta-data.

DETAILED DESCRIPTION OF THE INVENTION

The present invention will be described as implemented in a programmed digital computer. It will be understood that a person of ordinary skill in the art of digital image processing and software programming will be able to program a computer to practice the invention from the description given below. The present invention may be embodied in a computer program product having a computer readable storage medium such as a magnetic or optical storage medium bearing machine readable computer code. Alternatively, it will be understood that the present invention may be implemented in hardware or firmware.

Referring first to Fig. 1, a digital image processing system useful for practicing the present invention is shown. The system generally designated **10**, includes a digital image processing computer **12** connected to a network **14**. The digital image processing computer **12** can be, for example, a Sun Sparcstation, and the network **14** can be, for example, a local area network with sufficient capacity to handle large digital images. The system includes an image capture device **15**, such as a high resolution digital camera, or a conventional film

camera and a film digitizer, for supplying digital images to network **14**. A digital image store **16**, such as a magnetic or optical multi-disk memory, connected to network **14** is provided for storing the digital images to be processed by computer **12** according to the present invention. The system **10** also includes one or more display devices, such as a high resolution color monitor **18**, or hard copy output printer **20** such as a thermal or inkjet printer. An operator input, such as a keyboard and track ball **21**, may be provided on the system.

Referring next to Fig. 2, at least two source digital images are provided **200** in the method of the present invention. The source digital images can be provided by a variety of means; for example, they can be captured from a digital camera, extracted from frames of a video sequence, scanned from photographic film or hardcopy output, or generated by any other means. A cropping aspect ratio L:H is also provided **202**. The cropping aspect ratio is the ratio of the length (distance of the horizontal edge) to the height (distance of the vertical edge) of the desired cropped digital image. For example, in a digital photofinishing system that prints hardcopies of images that have been digitized from film after being captured by an Advanced Photo System (APS) camera, the cropping aspect ratio is constrained to be either 16:9, 3:2, or 3:1, corresponding to HDTV, Classic, and Panoramic formats, respectively.

A cropping criterion is also provided **204**. The cropping criterion specifies the size and location of the cropped digital image. In the preferred embodiment, the cropping criterion states that the cropped digital image be the composite digital image region having the largest area out of the set of all regions having aspect ratio L:H. In an alternative embodiment, the cropping criterion is that the cropped digital image be the composite digital image region having the largest area out of the set of all regions having aspect ratio L:H and having centers at the centroid of the composite digital image. In yet another alternative embodiment, the cropping criterion is that the cropped digital image be the composite digital image region having the largest area out of the set of all regions having aspect ratio L:H and having centers at the centroid of the main subject of the composite digital image.

The source digital images are then combined **206** by a scheme known in the art for combining images captured from the same nodal point, similar nodal points, or different nodal points, to form a composite digital image. In step **208**, a cropping region is selected, the cropping region being a composite digital image region having aspect ratio L:H provided in step **202**, selected according to the cropping criterion provided in step **204**. Once the cropping region has been selected **208**, the composite digital image is cropped **210** to the cropping region, yielding the cropped digital image **212**.

In one embodiment, the current invention further comprises the step of resizing **214** the cropped digital image. For example, consider the digital photofinishing system that prints hardcopies of images that have been digitized from film at an aspect ratio of 3:2, and requires the spatial resolution of images to be 6000 pixels by 4000 pixels. If four digital images are provided to the method of Fig. 2, each digital image having a spatial resolution of 6000 pixels by 4000 pixels, the cropped digital image may have spatial resolution 9000 pixels by 6000 pixels. In order to render a hardcopy print of the cropped digital image through the digital photofinishing system, the cropped digital image is resized to have spatial resolution 6000 pixels by 4000 pixels. The resizing step can be performed by any technique known in the art; for example, bilinear interpolation, bicubic interpolation, spline interpolation, or any of a variety of other image resizing techniques (see textbook: A.K. Jain, "Fundamentals of Digital Image Processing," Prentice Hall, 1989, Chapter 4, pp. 80-131, for a discourse on image sampling and resizing).

In another embodiment, the current invention further comprises the step of transforming **216** the pixel values of the cropped digital image to an output device compatible color space. The output device compatible color space can be chosen for any of a variety of output scenarios; for example, video display, photographic print, ink-jet print, or any other output device.

Referring next to Fig. 3, the preferred cropping criterion is illustrated. The source digital images **300** overlap in overlapping pixel regions **302**. In step **206**, the source digital images are combined to form the composite

digital image **304**. The cropping region **306** is then selected in step **208** according to the cropping criterion **204**. In the preferred embodiment, the cropping region **306** has the largest area of all composite digital image regions having aspect ratio L:H. In some instances, there can be more than one distinct composite digital image region having aspect ratio L:H and having maximum area, yielding multiple candidate regions for the cropping regions. In such instances, there may be a small (e.g. less than 10) or very large (e.g. more than 10) set of candidate regions. Furthermore, in instances where there is an very large set of candidate regions, the centroids of the candidate regions may form one or more path segments.

If only one candidate region exists, it is chosen as the cropping region. If a small number of candidate regions exist, the cropping region is chosen randomly from the small set of candidate regions. If a very large number of candidate regions exist, and the centroids of those candidate regions form a single path segment, the cropping region is chosen to be the candidate region whose center corresponds to the center of the path segment. If a very large number of candidate regions exist, and the centroids of those candidate regions form more than one distinct path segment, one path segment is chosen at random, and the cropping region is chosen to be the candidate region whose center corresponds to the center of that path segment.

Referring next to Fig. 4, another embodiment of the cropping criterion is illustrated. The cropping region **400** of the composite digital image **402** is the region having the largest area of all composite digital image regions having aspect ratio L:H, and having a center at the centroid **404** of the composite digital image **402**.

Referring next to Fig. 5, another embodiment of the cropping criterion is illustrated. The cropping region **500** of the composite digital image **502** is the region having the largest area of all composite digital image regions having aspect ratio L:H, and having a center at a main subject **504** of the composite digital image **502**. The main subject **504** of the composite digital image **502** can be indicated manually, or determined automatically by techniques

known in the art; see US Patent 6,282,317 issued August 28, 2001 to Luo et al., for an example of automatic main subject detection, which is incorporated herein by reference.

Referring next to Fig. 6, a block diagram of the method for choosing a candidate region according to the preferred cropping criterion **204** is shown. First, any composite digital image region having aspect ratio L:H and maximum area is identified **600** as a candidate region. A query **602** is made as to whether there is a single candidate region. An affirmative response means that the candidate region is chosen **604** as the cropping region. A negative response leads to a query **606** as to whether there are a small number of candidate regions. An affirmative response means that one candidate region **608** is chosen as the cropping region. The candidate region can be chosen arbitrarily, for example, by listing all of the candidate regions, and then by choosing the first candidate region in the list. A negative response leads to a query **610** as to whether there is a single path containing centers of candidate regions. An affirmative response leads to the choice **612** of the candidate region whose center corresponds to the midpoint of the path as the cropping region. A negative response leads to the choice **614** of one of the paths. The path can be chosen arbitrarily for example, by listing all of the paths, and then by choosing the first path in the list. The candidate region whose center corresponds to the midpoint of the chosen path is chosen **616** as the cropping region.

Referring next to Fig. 7, the step **200** of providing at least two source digital images further comprises the step **704** of applying a metric transform **702** to a source digital image **700** to yield a transformed source digital image **706**. A metric transform refers to a transformation that is applied to the pixel values of a source digital image, the transformation yielding transformed pixel values that are linearly or logarithmically related to scene intensity values. In instances where metric transforms are independent of the particular content of the scene, they are referred to as scene independent transforms.

In one example of such an embodiment, a source digital image **700** was provided from a digital camera, and contains pixel values in the sRGB color

space (see Stokes et al., "A Standard Default Color Space for the Internet - sRGB", http://www.color.org/sRGB.html, pp.1-12). A metric transform **702** is used to convert the pixel values into nonlinearly encoded Extended Reference Input Medium Metric (ERIMM) (PIMA standard #7466, found on the World Wide Web at (http://www.pima.net/standards/it10/IT10_POW.htm), so that the pixel values are logarithmically related to scene intensity values.

The metric transform is applied to rendered digital images, i.e. digital images that have been processed to produce a pleasing result when viewed on an output device such as a CRT monitor or a reflection print. For digital images encoded in the sRGB metric transform is a gamma compensation lookup table that is applied to the source digital image **700** first. The formula for the gamma compensation lookup table is as follows. For each code value *cv*, ranging from 0 to 255, an exposure value *ev* is calculated based on the logic:

$$\text{if } (cv <= 10.015) \qquad ev = cv / (255 * 12.92)$$

otherwise

$$ev = (cv / 255) + 0.055) / 1.055)^{0.45}$$

Once the pixel values are modified with the gamma compensation lookup table, a color matrix transform is applied to compensate for the differences between the sRGB color primaries and the ERIMM metric color primaries. The nine elements of the color matrix τ are given by:

$$\begin{matrix} 0.5229 & 0.3467 & 0.1301 \\ 0.0892 & 0.8627 & 0.0482 \\ 0.0177 & 0.1094 & 0.8727 \end{matrix}$$

The color matrix is applied to the red, green, blue pixel data as

$$R' = \tau_{11} R + \tau_{12} G + \tau_{13} B$$

$$G' = \tau_{21} R + \tau_{22} G + \tau_{23} B$$

$$B' = \tau_{31}\,R + \tau_{3}\,G + \tau_{33}\,B$$

where the R, G, B terms represent the red, green, blue pixel values to be processed by the color matrix and the R', G', B' terms represent the transformed red, green, blue pixel values. The R', G', and B' pixel values are then converted to a log domain representation thus completing the metric transformation from sRGB to ERIMM.

Referring next to Fig. 8A, the step **200** of providing at least two source digital images further comprises the step **804** of modifying the pixel values of at least one of the source digital images **800** by a linear exposure transform so that the pixel values in the overlap regions of overlapping source digital images are similar. A linear exposure transform refers to a transformation that is applied to the pixel values of a source digital image, the transformation being linear with respect to the scene intensity values at each pixel. Examples of linear exposure transforms can be found in the aforementioned Cahill, Gindele, Gallagher, and Spaulding reference.

Referring next to Fig. 8B, the step **200** of providing at least two source digital images further comprises the step **802** of modifying the pixel values of at least one of the source digital images **800** by a radial exposure transform so that any light falloff present in the source digital images is compensated. A radial exposure transform refers to a transformation that is applied to the pixel values of a source digital image, the transformation being a function of the distance from the pixel to the center of the image. Examples of radial exposure transforms can be found in the aforementioned Cahill and Gindele reference.

Referring next to Fig. 9, a more detailed description of the step **206** of combining source digital images is described. The source digital images are geometrically warped **900** to compensate for distortion due to perspective projection. In a physical sense, this distortion would not exist if the sensor were not planar, but rather spherical (with the radius of the sphere depending on the focal length of the lens). The warped source digital images are then aligned **902** to identify the overlapping regions. The alignment procedure is performed using

any of the aforementioned techniques, such as phase correlation or cross correlation. Once the source digital images have been aligned, they are blended **904** in the overlapping regions.

Referring next to Fig. 10, the blending step **904** further comprises a feathering scheme, weighted averages, or some other technique known in the art, to form a composite digital image. In one embodiment, a pixel **1002** in the overlap region **1004** is assigned a value based on a weighted average of the pixel values from both source digital images **1000**; the weights are based on its relative distances **1006** to the edges of the source digital images **1000**.

Referring next to Fig. 11, a further embodiment of the step **904** of blending warped digital images is described. The warped digital images are projected **1100** to simulate capture on parallel image planes. This is done by estimating the fundamental matrix relating the two images. The fundamental matrix, described in the aforementioned Zhang, Deriche, Faugeras, and Luong reference, contains all of the information pertinent to the geometrical relationship between two cameras. Once the warped digital images have been projected **1100** to simulate capture on parallel image planes, they are morphed **1102** using a standard image morphing procedure, such as the procedure described in the aforementioned Beier and Neely reference, producing a morphed digital image. A set of viewing parameters at which to view the morphed digital image is chosen **1104**, and the morphed digital image is then reprojected **1106** to the chosen viewing parameters, producing the composite image. The blending step **904** described in this embodiment is the well known view morphing procedure, described in the aforementioned Seitz and Dyer reference.

Referring next to Figs. 12A and 12B, the aspect ratio of an image **1200** is defined as the ratio of the length **1202** of the image to its height **1204**. When the width **1204** of the image is greater than its length **1202**, as depicted in Fig. 12A, the aspect ratio is less than one, and is referred to as a portrait aspect ratio. When the width **1204** of the image is less than its length **1202**, as depicted in Fig. 12B, the aspect ratio is greater than one, and is referred to as a landscape aspect ratio. Advanced Photo System (APS) cameras provide the choice of three

different aspect ratios: HDTV (H), a 16:9 aspect ratio, Classic (C), a 3:2 aspect ratio, or Panoramic (P), a 3:1 aspect ratio. These aspect ratios are all landscape aspect ratios, but the APS camera can be rotated to capture images with the corresponding portrait aspect ratios 9:16, 2:3, and 1:3.

Referring next to Figs. 13A and 13B, at least one of the source digital image files **1300** may contain meta-data **1304** in addition to the image data **1302**. Such meta-data can include the cropping aspect ratio **1306**, or any information pertinent to the pedigree of the source digital image.

The invention has been described in detail with particular reference to certain preferred embodiments thereof, but it will be understood that variations and modifications can be effected within the spirit and scope of the invention.

PARTS LIST

10	digital image processing system
12	digital image processing computer
14	network
15	image capture device
16	digital image store
18	high resolution color monitor
20	hard copy output printer
21	keyboard and trackball
200	provide source digital images step
202	provide cropping aspect ratio step
204	provide cropping criterion step
206	combine source digital images step
208	select cropping region step
210	crop composite digital image step
212	cropped digital image
214	resize cropped digital image step
216	transform pixel values step
300	source digital images
302	overlapping pixel regions
304	composite digital image
306	cropping region
400	cropping region
402	composite digital image
404	centroid of composite digital image
500	cropping region
502	composite digital image
504	main subject of composite digital image
600	identify candidate regions step
602	single candidate region query
604	cropping region chosen
606	finite number of candidate regions query
608	cropping region chosen
610	single path of candidate region centers query

612	cropping region chosen
614	choose first path step
616	cropping region chosen
700	source digital image
702	metric transform
704	apply metric transform step
706	transformed source digital image
800	source digital images
802	modify with linear exposure transform step
804	modify with radial exposure transform step
900	warp source digital images step
902	align warped digital images step
904	blend warped digital images step
1000	source digital images
1002	pixel
1004	overlapping pixel region
1006	distances to image edges
1100	project warped digital images step
1102	morph projected digital images step
1104	choose viewing parameters step
1106	re-project morphed digital image step
1200	image
1202	length
1204	width
1300	source digital image file
1302	image data
1304	meta-data
1306	aspect ratio